Title
Distentangling the systems contributing to changes in learning during adolescence.

Authors
Master, Sarah L
Eckstein, Maria K
Gotlieb, Neta
et al.

Publication Date
2020-02-01

DOI
10.1016/j.dcn.2019.100732

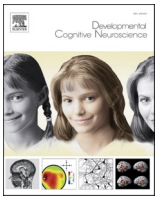
Peer reviewed

Developmental Cognitive Neuroscience 41 (2020) 100732

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: www.elsevier.com/locate/dcn



Distentangling the systems contributing to changes in learning during adolescence

Sarah L. Master [a], Maria K. Eckstein [a], Neta Gotlieb [a], Ronald Dahl [c], Linda Wilbrecht [a,b], Anne G. E. Collins [a,b]

[a] Department of Psychology, University of California, Berkeley, United States
[b] Helen Wills Neuroscience Institute, University of California, Berkeley, United States
[c] Institute of Human Development and School of Public Health, University of California, Berkeley, United States

ARTICLE INFO

Keywords:
Development
Reinforcement learning
Working memory
Computational modeling
Adolescence

ABSTRACT

Multiple neurocognitive systems contribute simultaneously to learning. For example, dopamine and basal ganglia (BG) systems are thought to support reinforcement learning (RL) by incrementally updating the value of choices, while the prefrontal cortex (PFC) contributes different computations, such as actively maintaining precise information in working memory (WM). It is commonly thought that WM and PFC show more protracted development than RL and BG systems, yet their contributions are rarely assessed in tandem. Here, we used a simple learning task to test how RL and WM contribute to changes in learning across adolescence. We tested 187 subjects ages 8 to 17 and 53 adults (25-30). Participants learned stimulus-action associations from feedback; the learning load was varied to be within or exceed WM capacity. Participants age 8-12 learned slower than participants age 13-17, and were more sensitive to load. We used computational modeling to estimate subjects' use of WM and RL processes. Surprisingly, we found more protracted changes in RL than WM during development. RL learning rate increased with age until age 18 and WM parameters showed more subtle, gender- and puberty-dependent changes early in adolescence. These results can inform education and intervention strategies based on the developmental science of learning.

1. Introduction

There is increasing evidence that multiple neural systems contribute to human learning (Hazy et al., 2007; Myers et al., 2002), even in simple cognitive paradigms previously modeled with a single learning process (Bornstein & Daw, 2012; Bornstein & Norman, 2017; Collins & Frank, 2012). Basal ganglia (BG) dependent reinforcement learning (RL) processes are thought to be supplemented by multiple other systems, including prefrontal executive functions (Badre et al., 2010) such as working memory (WM; (Collins & Frank, 2012)) and model-based planning (Daw et al., 2011), as well as hippocampus-based episodic memory (Bornstein & Daw, 2012; Bornstein & Norman, 2017; Davidow et al., 2016; Myers et al., 2002; Wimmer et al., 2014). To understand developmental changes in learning, it is important to carefully capture the contributions of these multiple systems to learning. Previous work has shown differential developmental trajectories for RL, episodic memory, and model-based planning (Crone et al., 2006; Decker et al., 2016; Selmeczy et al., 2018; Somerville et al., 2010). Here, we investigate how the relative contributions of RL and WM change during development.

RL is an incremental learning process which updates stored choice values from the discrepancy between obtained and expected reward (the reward prediction error; RPE) in proportion to a learning rate, in order to maximize future rewards (Sutton & Barto, 2017). This process is thought to be implemented via dopamine-dependent plasticity in cortico-striatal circuits. It has been proposed that the BG are 'mature' by mid adolescence, but it is also known that structures that provide inputs to the BG continue to show anatomical and functional change (Braams et al., 2015; Casey et al., 2008; Galvan et al., 2006a; Galvan et al., 2019; Van Den Bos et al., 2012). Sensitivity to rewarding outcomes in the ventral striatum has been shown to peak during mid-adolescence, relative to children and adults (Braams et al., 2015; Galvan et al., 2006a; Gulley et al., 2018; Hauser et al., 2015; Somerville et al., 2010). Mid-teenage adolescents have also been shown to prioritize positive feedback and neglect negative feedback more than adults when learning from reinforcement (Cohen et al., 2010; Davidow et al., 2016; Jones et al., 2014; Palminteri et al., 2016; van den Bos et al., 2012), and 13 to 17-year-olds have been shown to use lower learning rates than adults (Davidow et al., 2016; Jones et al., 2014). However, work on the development of RL is heterogeneous and appears dependent on specific task characteristics (van den Bos et al., 2009). We explore if BG-dependent RL continues to develop throughout adolescence into

https://doi.org/10.1016/j.dcn.2019.100732
Received 29 April 2019; Received in revised form 23 September 2019; Accepted 4 November 2019
Available online 14 November 2019

adulthood and examine the possibility that developmental changes in learning are driven by different systems at different times, such as development of WM, in the context of a simple conditional associative learning task.

WM allows for the immediate and accurate storage of information, but representations in WM are thought to decay quickly with time and are subject to interference, as the information that can be held in WM is limited (D'Esposito & Postle, 2015; Oberauer et al., 2018). Therefore there may be tradeoffs to using fast capacity-limited WM, versus slower, capacity unlimited RL-based learning systems. The prefrontal cortex (PFC) is known to develop late into adolescence and the third decade of life (Casey et al., 2008; Giedd et al., 1999; Larsen & Luna, 2018) and is thought to be critical for WM performance (Curtis & D'Esposito, 2003; Miller & Cohen, 2001). The use of WM in complex tasks has been shown to improve during development into late adolescence (Geier et al., 2009; Huizinga et al., 2006; Luna, 2009; McAuley & White, 2011), though there is some evidence that the use of WM in simple tasks develops in early childhood (Crone et al., 2006; Potter et al., 2017).

Behavioral testing and computational modeling can be used to disentangle the simultaneous contributions of RL and WM in human learning, and how their use differs between individuals. Behavior on classic one-step reward learning tasks is typically modeled with single-process RL models. However, the use of WM for short-term storage is an important component of human learning. Using a deterministic reward-learning task called "RLWM" that taxes WM by varying the amount of information to learn in each block, we have previously isolated contributions of WM and RL learning (Collins, 2018; Collins & Frank, 2012,2018). In multiple studies we found that participants mainly used WM for learning when the load was within their capacity, and otherwise compensated with RL. We also found that learning deficits in schizophrenia were a result of weakened WM contributions with intact RL (Collins & Frank, 2014; Collins et al., 2017). By accounting for WM in our task and model, we can extract the unconfounded separate contributions of both WM and RL. Here we used the same approach to investigate the maturation of WM and RL and their relative contribution to learning across adolescent development (sampling subjects 8-17 and 25-30).

Using behavioral testing and computational modeling, we examined three separate hypotheses of how RL- and WM-based learning develop relative to each other. Our first hypothesis was that both RL and WM systems' contributions to learning would show protracted development into later adolescence (age 17) such that both systems are dynamic throughout the pre-teenage and teenage years (8-17). Developmental changes in WM are prominent in the literature on development of executive function, including the maintenance (Geier et al., 2009; McAuley & White, 2011) and manipulation of information in WM (Crone et al., 2006; Huizinga et al., 2006), as well as the precision of the representations in WM (Luna, 2009). There is also a strong literature showing changes in RL learning systems from childhood to adulthood, such as dynamic changes in reward sensitivity in the striatum across adolescence (Braams et al., 2015; Cohen et al., 2010; Davidow et al., 2016; Somerville et al., 2010) and changes in learning rate between adolescence and adulthood (Davidow et al., 2016).

Our second hypothesis emphasized the relative importance of WM development over that of RL in accounting for changes in learning in adolescence. Dual systems models and other popular models of adolescent development place great weight on the late maturation of the PFC and PFC-dependent executive functions, such as WM or model-based learning (Casey et al., 2008; Decker et al., 2016; Huizinga et al., 2006; Steinberg, 2005). Functional activation of parietal cortex, also involved in WM and attention, has been shown to develop from ages 9 to 18 and to correlate with visuospatial WM performance (Klingberg et al., 2002). Additionally, the communication of PFC and parietal cortex through the fronto-parietal WM network is thought to increase from ages 8 to 18 (Klingberg, 2006; Nagy et al., 2004). Our second hypothesis therefore predicts that even though RL may be developing, we should observe more protracted development of WM systems and/or stronger effects of age (in the range 8-17) on WM. If true then we might conclude that WM changes are the primary drivers of changes in learning through late adolescence.

Finally, we hypothesized that pubertal onset may significantly impact WM processes. There is growing evidence that gonadal hormones affect inhibitory neurotransmission and other variables in the prefrontal cortex of rodents (Delevich et al., 2019a; 2019b; Juraska & Willing, 2017; Piekarski et al., 2017a). We therefore predicted that WM parameters would differ in children with different pubertal status or gonadal hormone concentration.

To evaluate these hypotheses, we tested children and adolescents aged 8 to 17 years old and adults aged 25 to 30 years old on the RLWM task (Collins & Frank, 2012). We then fit computational models of behavior to subjects' performance and assessed how these parameters changed with age, pubertal development and salivary testosterone levels. Using these established methods to disentangle the contributions of RL and WM, we found changes in RL contributions spanning adolescent development, but much weaker changes in WM contributions. WM differences did show relationships with pubertal variables.

Overall, these data support the somewhat surprising conclusion that changes in RL systems are important drivers of change in simple associative learning throughout adolescence. The results also support further inquiry into the role of pubertal processes in WM function in early adolescence.

2. Methods

2.1. Subject testing

All procedures were approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley. After entering the testing room, subjects under 18 years old and their guardians provided their informed assent or permission. All guardians were asked to fill out a demographic form. Subjects were led into a quiet testing room in view of their guardians, where they used a video game controller to complete four computerized tasks. An hour after the start of the experimental session and in between tasks, subjects provided a 1.8 mL saliva sample. At the conclusion of the tasks, subjects were asked to complete a short questionnaire which assessed their pubertal development (Petersen & Crockett, 1988) and collected other basic information, like their height and weight. For subjects under 10 years old, guardians completed the pubertal development questionnaire on behalf of their children. Subjects were then compensated with one $25 Amazon gift card for their time.

Participants over 18 provided informed consent and completed all forms themselves. They also answered retroactive questions about puberty, otherwise all testing procedures were identical.

2.2. Exclusion criteria

Potential subjects or their guardians were called prior to the experimental session to complete a verbal pre-screening questionnaire. Potential subjects were required to have normal or corrected-to-normal vision, and to be fluent in English. They could not have any previous or current neurological or psychological conditions, history of head injury or concussion, be on any prescription medications, or be colorblind.

2.3. Demographics

We recruited 191 children and 55 adults to participate in this study. Out of those who reported their race, 60 subjects identified as Asian, 10 African-American, and 6 Native American or Pacific Islander. 28 were of mixed race. The remaining 127 subjects identified as Caucasian. 29 subjects identified as Hispanic. 4 children and 1 adult failed to complete the task, either out of disinterest or because of controller malfunction.

All subjects who completed the task performed above chance (33% accuracy), so none were excluded for poor task performance. 187 children (89 female, mean (std) age 12.62 (2.76) years) and 54 adults (28 female, mean (std) age 26.77 (1.49) years) were included in analyses.

2.4. Experimental paradigm

The experiment described in this work was the second of four administered tasks. The preceding task was a 5 to 10 minute deterministic learning task in which subjects learned to select one correct action out of four possible actions. Halfway through, the correct action switched. The results for this and the other tasks will be reported elsewhere.

This experiment is based on the "RLWM" task described in (Collins et al., 2017; Collins & Frank, 2012, 2018), which was adapted for the developmental population. To make the task more engaging for children, all participants were told that aliens had landed on earth and wanted to teach us their alien language. In this alien language, each image on screen was matched with one button on the controller. Participants completed one block of training and then ten independent learning blocks, for a total duration of less than 25 minutes (mean duration 16.9 minutes, range 14 – 25 minutes).

In each block, subjects were presented with a new set of visual stimuli of set size *ns*, with set sizes between $ns = 2$ and $ns = 5$. Each visual stimulus was presented 12-14 times in a pseudo-randomly interleaved manner (controlling for a uniform distribution of delay between two successive presentations of the same stimulus within [1:2**ns*] trials, and the number of presentations of each stimulus), for a total of *ns*13 trials. At each trial, a stimulus was presented centrally on a black background (Fig. 1). Subjects had up to 7 seconds to answer by pressing one of three buttons on the controller. Key press was followed by visual feedback presentation for 0.75 seconds, then a fixation period of 0.5 seconds before the onset of the next trial. For each image, the correct key press was the same for the entire block, and all feedback was truthful. Upon pressing the correct key subjects were shown "Correct" while pressing any other key led to "Try again!" Failure to answer within 7 seconds was indicated by a "No valid answer" message. Stimuli in a given block were all from a single category of familiar, easily identifiable images (e.g. colors, fruits, animals) and did not repeat across blocks. Participants were shown all stimuli at the beginning of each block, and encouraged to familiarize themselves with them prior to starting the learning block.

This task engages reinforcement learning through the repetition of simple actions which lead to reward. Working memory is a capacity-limited resource which can be leveraged to quickly improve accuracy when small sets of stimulus-action pairings must be learned. To tax working memory more or less across the task and thus assess WM contributions to learning, we varied the set size across blocks: out of 10 blocks, 3 had set size $ns = 2$, 3 set size $ns = 3$, 2 set size $ns = 4$, and 2 set size $ns=5$. We have shown in previous work that varying set size provides a way to investigate the contributions of capacity- and resource-limited working memory to reinforcement learning.

2.5. Saliva collection

In addition to self-report measures of pubertal development, we also collected saliva from each of our subjects to quantify salivary testosterone. Testosterone is a reliable measure of pubertal status in boys and girls and is associated with changes in brain and cognition in adolescence (Herting et al., 2014; Peper et al., 2011).

Subjects refrained from eating, drinking, or chewing anything at least an hour before saliva collection. Subjects were asked to rinse their mouth out with water approximately 15 minutes into the session. At least one hour into the testing session, they were asked to provide 1.8 mL of saliva through a plastic straw into a 2 mL tube. Subjects were instructed to limit air bubbles in the sample by passively drooling into the tube, not spitting. Subjects were allotted 15 minutes to provide the sample. After the subjects provided 1.8 mL of saliva, or 15 minutes had passed, the sample was immediately stored in a −20 °F freezer. The date and time were noted by the experimenter. The subjects then filled out a questionnaire of information which might affect the hormone concentrations measured in the sample (i.e. whether the subjects had recently exercised).

All subjects were asked whether they would like to complete two more saliva samples at home for additional compensation (another $25 Amazon gift card). Subjects who agreed to do the optional saliva samples were sent home with two 2 mL tubes, straws, and questionnaires identical to the one completed in lab. They were asked to complete each sample between 7:00 and 9:00 am on two different days following the

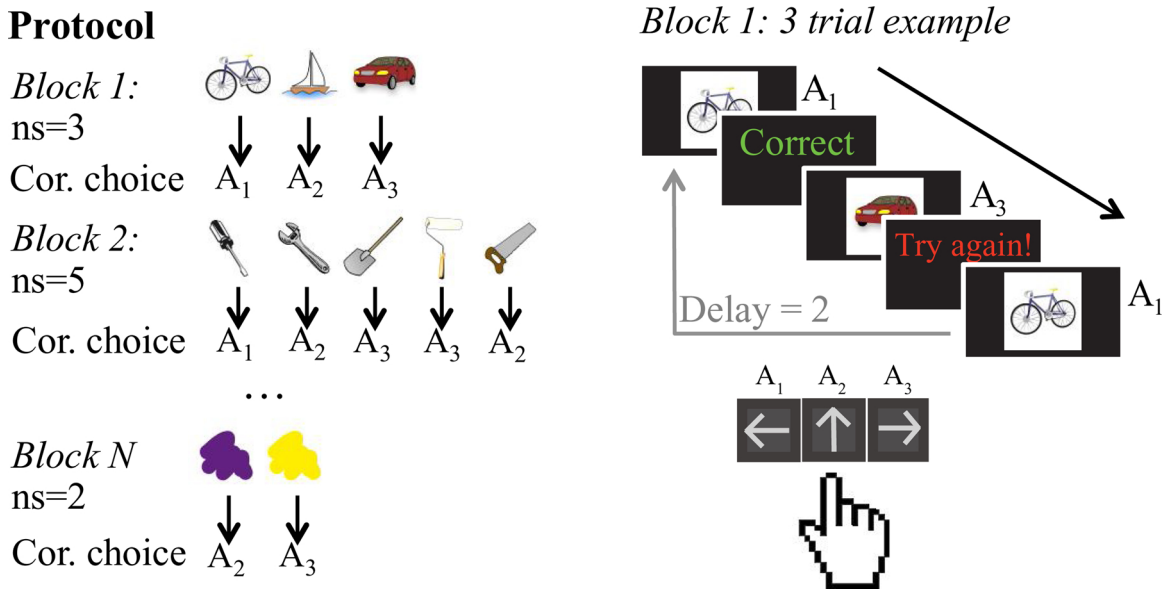


Fig. 1. Experimental protocol. In each block, participants learned to select the correct action for each image for a new set of images. At the beginning of each block, the full set of images was presented for familiarization, then single trials began. On each trial of a block, participants responded to each stimulus by pressing one of three buttons on a hand-held controller. Immediately after responding they received deterministic truthful feedback (Correct/Try again!), before moving on to the next trial after a fixed interval. Crucially, participants needed to learn different numbers of stimuli (set size) in different blocks. Set size (ns) varied from 2 to 5.

testing session. Subjects were asked to refrain from eating, drinking, or brushing their teeth before doing the sample, and to fill out each questionnaire as soon as they were finished collecting saliva, taking care to note the date and time of the sample. Subjects were also instructed to keep the samples in the freezer, then wrap them in a provided ice pack in order to deliver the samples to the lab. Once both samples were complete, subjects contacted an experimenter and scheduled a time to return the samples, who gave the subjects their additional compensation, took note of any abnormalities in the samples, and immediately stored them in a -20 degree freezer. Samples in the -20 degree freezer were transferred weekly to a -80 degree freezer in an adjacent facility.

2.6. Salivary Testosterone Testing

Salivary testosterone was quantified using the Salimetrics Salivatory Testosterone ELISA (cat. no. 1-2402, Bethesda, MA). Intra- and inter-assay variability for testosterone were 3.9% and 3.4%, respectively. Samples below the detectable range of the assay were assigned a value of 5 pg/mL, 1 pg below the lowest detectable value. Final testosterone sample concentration data were cleaned with a method developed by Shirtcliff and Byrne (in prep). Specifically, we produced a mean testosterone concentration from every salivary sample obtained from each of our subjects (every subject provided 1 to 3 samples). Subjects who had multiple samples below the detectable range of the assay (6 pg/mL) had their mean testosterone concentration replaced with 1 pg below the lowest detectable value (5). There were no subjects with any samples above the detectable range. Within subjects aged 8 to 17 only, outliers greater than 3 standard deviations above the group mean were fixed to that value, then incremented in values of +0.01 to retain the ordinality of the outliers.

2.7. Model-independent analyses

We now describe how we analyzed the data from the behavioral task. To describe our data at a high level and illustrate developmental trends in performance, we first analyzed overall accuracy as a function of precise subject age. To assess learning, we calculated the proportion of correct trials for each subject on each stimulus iteration. Each individual stimulus was repeated 12 to 14 times within a block. Within each set size, we calculated each subject's average percentage of correct responses at each stimulus iteration (Fig. 2a).

We quantified the effect of set size on performance by calculating a set size slope (Fig. 2b). The set size slope was a linear contrast of the form:

$$-1.5 * perf(ns5) - 0.5 * perf(ns4) + 0.5 * perf(ns3) + 1.5 * perf(ns2)$$

where perf(ns) is the average overall performance of trials within a block of set size ns. We then analyzed set size slope as a function of age.

In order to more precisely assess the effects of our set size manipulation on performance, we used a logistic regression to model trial-by-trial accuracy as a function of previous correct trials (pcor), previous incorrect trials (pinc), number of trials since the last presentation of the same stimulus (delay), and set size (ns) as predictors (Collins & Frank, 2012).

To understand the effects of development on behavior and on model parameters, we ran a series of analyses at group and individual levels. First, following the practice in the literature to separate "children" from "adolescents" and adults (Potter et al., 2017) we grouped 8 to 12, 13 to 17, and 25 to 30 year-olds into separate groups and ran one-way ANOVAs on behavioral and modeling measures of interest, looking for broad group-based, age-related differences in cognition; ANOVAs were replaced by Kruskal-Wallis non-parametric tests for non-normal

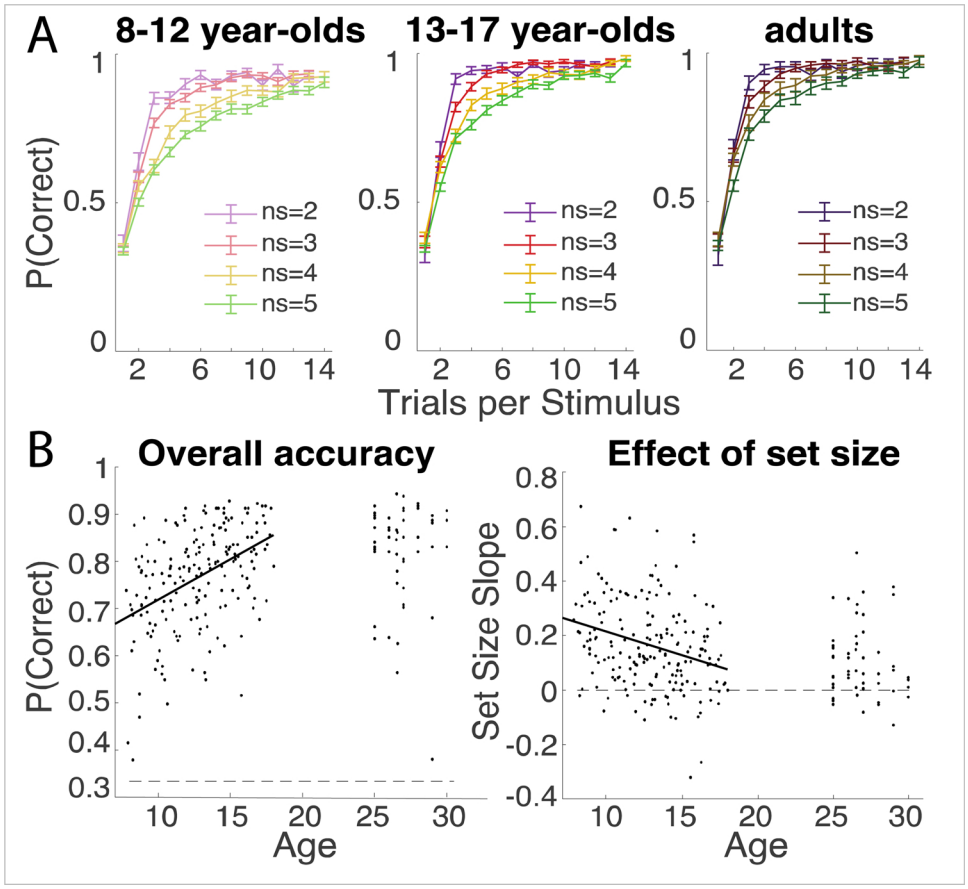


Fig. 2. Age effects on behavior. A. Learning curves indicate the proportion of correct trials as a function of the number of encounters with given stimuli by set size (ns) in subjects aged 8 to 12, 13 to 17, and 25 to 30. All subjects quickly reached asymptotic accuracy in set sizes 2 and 3. In set sizes 4 and 5, learning was more graded. 8-12-year-olds appeared to learn slower than both 13-17-year-olds and adults (25-30). 13-17-year-olds and 25-30-year-olds exhibited similar learning. B. Overall accuracy and set size slope (linear effect of set size on performance; see methods) as a function of age. Each dot represents a single subject; lines indicate the best fit in a regression model. In subjects aged 8 to 17 overall accuracy significantly increased with age while set size effect significantly decreased, though it remained positive in all age groups.

measurements. Where group effects were present, we ran post-hoc tests to identify which group drove the effect. We used t-tests when the measure was normally distributed, rank tests otherwise.

While helpful to visualize results, the binning into coarse 8 to 12 and 13 to 17-year-old groups for non-adult participants was arbitrary. We also analyzed the data of this 8 to 17-year-old subsample as a function of age as a continuous measure. Specifically, we ran non-parametric (Spearman) correlation tests on behavioral measures with age as a continuous predictor within the non-adult group.

Additionally, to further investigate the effects of age and pubertal development on our outcome measures, we examined relationships between behavior and model parameters as a function of pubertal development score (PDS) and salivary testosterone, either as continuous predictors, or defining binned groups (see supplementary material). Given that girls tend to begin puberty earlier than their male peers, and that the production of testosterone tracks the development of boys and girls differently, we also ran analyses of pubertal effects separating male and female subjects, and combining them. We also grouped participants aged 8 to 17 into narrower age bins based on quartiles. Further descriptions of how we grouped subjects into age, PDS, and testosterone bins, plus additional statistical test methods and results, are included in the supplemental material.

3. Computational modeling

We used computational modeling to fit subject behavior and better quantify the separate involvements of reinforcement learning and working memory in task execution. We have shown that our model is able to separate the contributions of these two learning systems in both general (Collins, 2018; Collins & Frank, 2012, 2018) and specific populations (Gold et al., 2017). We tested 6 candidate models all built upon a simple reinforcement learning (RL) algorithm.

3.1. Classic RL

The simplest model is a two parameter Q-learner (*RL*), which updates the learned value Q for the selected action a given the stimulus s upon observing each trial's reward outcome r_t (1 for correct, 0 for incorrect):

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha\delta_t$$

where $\delta_t = r_t - Q_t(s,a)$ is the prediction error, and α is the learning rate, which is a free parameter. This model is similar to standard RL models as described in Sutton and Barto's Reinforcement Learning: An Introduction (2017). Choices are generated probabilistically with greater likelihood of selecting actions that have higher Q-values. This choice is driven by a softmax choice policy, which defines the rule for choosing actions in response to a stimulus:

$$p(a|s) = exp(\beta Q(s,a))/\Sigma_i(exp(\beta Q(s,a)))$$

Here, β is the inverse temperature parameter determining the degree to which differences in Q-values are translated into more deterministic choice, and the sum is over the three possible actions a_i. In this model, and all further models, all Q-values were initialized to $1/n_A$.

3.2. RL with undirected noise (RLe)

While the softmax allows for some stochasticity in choice, we also tested a model which allowed for "slips" of action. This was captured in an undirected noise parameter, ε. Given a model's policy $\pi = p(a|s)$, adding undirected noise consists in defining the new mixture choice policy:

$$\pi' = (1-\varepsilon)\pi + \varepsilon U$$

where U is the uniform random policy ($U(a) = 1/n_A$, with number of actions $n_A = 3$). ε is a free parameter constrained to values between 0 and 1. This undirected noise captures a choice policy where with probability $1 - \varepsilon$ the agent chooses an action based on the softmax probability, and with probability ε lapses and chooses randomly. Failing to account for this irreducible noise can allow model fits to be unduly influenced by rare odd data points, like those that may arise from attentional lapses (Nassar & Frank, 2016).

3.3. RL with positive learning bias (RLb)

To allow for potential neglect of negative feedback and bias towards positive feedback, we estimate a positive learning bias parameter *bias* such that for negative prediction errors ($\delta < 0$), the learning rate α is reduced by $\alpha = (1 - bias)\alpha$. Thus, values of *bias* near 1 indicate complete neglect of negative feedback, whereas values near 0 indicate equal learning from negative and positive feedback. We chose not to implement a separate α for positive and negative feedback, as has been done in modeling similar RL tasks (Frank et al., 2007; Katahira, 2015; Lefebvre et al., 2017). In previous work with this model we included both a positive and negative α, but consistently found a bias towards learning from positive feedback; thus parameterizing it as a bias was more efficient. Furthermore, this parameterization allows the *bias* parameter to be shared between RL and WM modules in the RLWM model, which increases model identifiability.

3.4. RL with forgetting (RLf)

In this model we allow for potential forgetting of Q-values on each trial, implemented as a decay at each trial toward the initial, uninformed Q-value Q_0:

$$Q_{t+1} = Q_t + \varphi(Q_0 - Q_t)$$

where $0 < \varphi < 1$ is the forgetting parameter and $Q_0 = 1/n_A$.

3.5. RL with 4 learning rates (RL4)

To improve the fit within the "RL only" class of models, we tested a version of the Q-learner that included a different learning rate α for each set size. Theoretically, this model could capture set size effects if they were driven by slower learning in higher set sizes.

3.6. RL and working memory (RLWM)

This model incorporates two separate mechanisms by which learning can take place which interact at the level of choice. The first mechanism is a RL model as described above, with an inverse temperature parameter β, learning rate α, positive learning bias parameter *bias*, and undirected noise ε. The second mechanism is a working memory module.

The WM module stores weights between stimuli and actions, *W(s,a)*, which are initialized similarly to RL Q-values. To model fast storage of information, we assume $W_{t+1}(s_t,a_t) = W_t(s_t,a_t) + \alpha_{WM}(r_t - W_t(s_t,a_t))$. With working-memory learning rate $\alpha_{WM} = 1$, this formula captures perfect retention of the previous trial's information, such that $W_{t+1}(s_t, a_t) = r_t$. The *bias* parameter is applied to the learning rate for both the RL and WM modules, and is thus a joint parameter, capturing failure to take into account negative feedback in both RL and WM systems. To model delay-sensitive aspects of working memory (where active maintenance is increasingly likely to fail with intervening time and other stimuli), we assume that WM weights decay at each trial according to $W_{t+1} = W_t + \varphi_{WM}(W_0 - W_t)$. The WM policy uses these weights in a softmax choice with added undirected noise, using the same noise parameters as the RL module.

RL and WM involvement in choice is modeled with a WM weight parameterized by the working memory capacity parameter K, and a WM confidence prior ρ. The overall choice policy is defined as a mixture

using WM weight $W_{WM} = \rho(\min(1,K/ns))$:

$$P(a|s) = W_{WM}P_{WM}(a|s) + (1 - W_{WM})P_{RL}(a|s)$$

ρ captures the subject's overall propensity to use WM vs. RL when within WM's capacity. The WM weight then considers the capacity limit of the WM module as indicated by the proportion of items that can be maintained in working memory (min(1,K/ns)), and thus can contribute to the policy, as well as the subjects' prior for relying on WM.

Note that our model assumes that information stored for each stimulus in working memory pertains to action-outcome associations. Furthermore, we approximate working memory by focusing on three key characteristics: 1) rapid and accurate encoding of information when

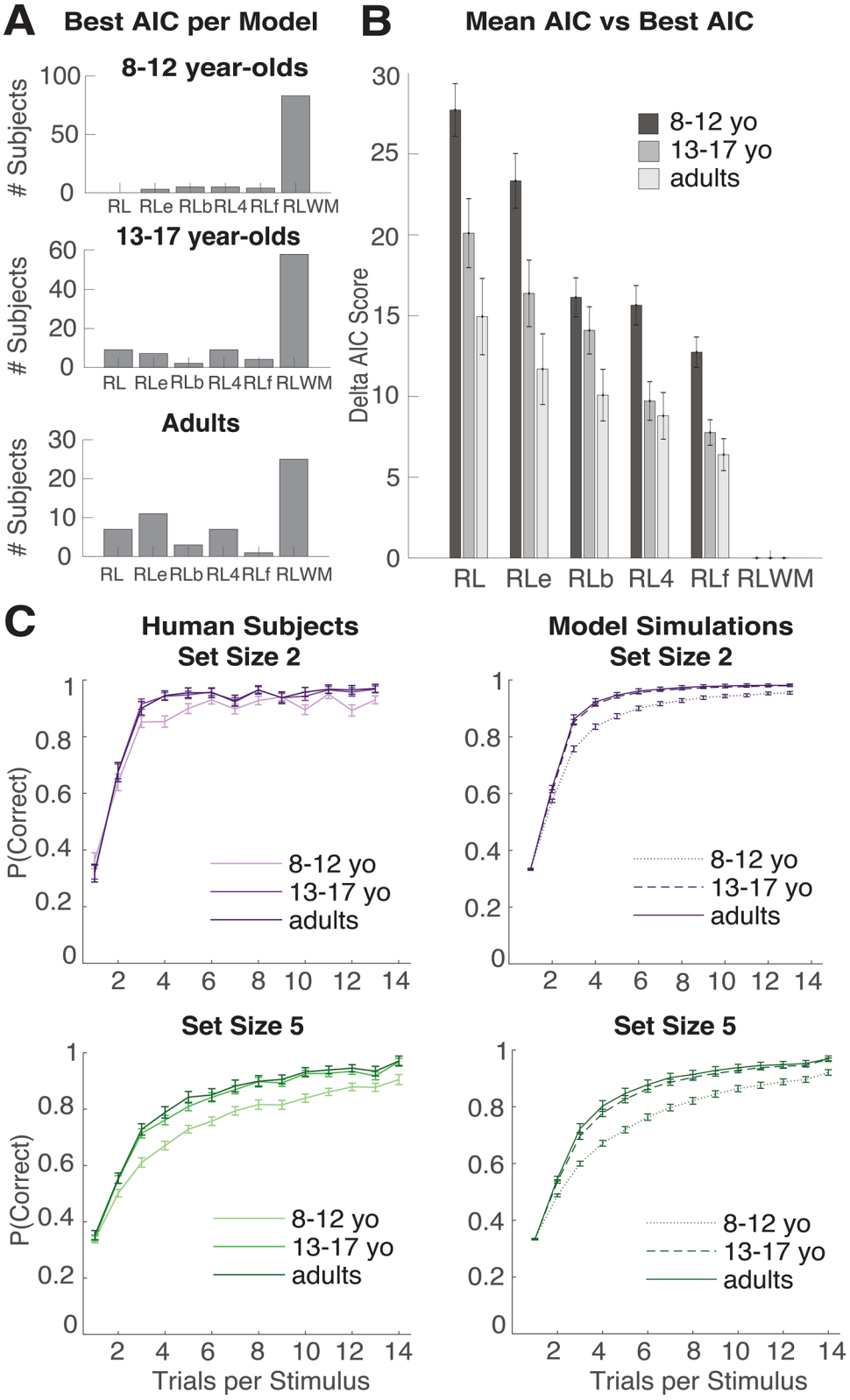


Fig. 3. Model validation. A. Best fit model per subject. All subjects' behavior was fit with 6 candidate models: reinforcement learning (RL), RL with an epsilon noise parameter (RLe), RL with perseveration (RLp), RL with four learning rates (RL4), RL with forgetting (RLf), and RL with working memory (RLWM; see methods). Plotted here is the number of subjects best fit by each candidate model in each age group as measured with AIC score. RLWM was the best-fitting model for a majority of subjects within subjects aged 8-12, 13-17, and 25-30. B. Difference in mean AIC score from the best fitting model (RLWM). Within subjects 8 to 12 years old (yo), 13 to 17 years old, and adults (25 to 30 yo), we calculated the mean AIC score for each candidate model, then compared to the mean AIC for the winning model (RLWM). Lower numbers indicate better fits. Error bars made with standard error of the mean. C. Model validation. Learning curves for participants (left) and model simulations (right) for set sizes 2 and 5 (see supplementary materials for all set sizes). RLWM model simulations with individual fit parameters accounted for behavior.

low amounts of information are to be stored; 2) decrease in the likelihood of accessing items from working memory when more information is presented than can be stored in its limited capacity or resource; 3) decay due to forgetting.

The strongest model in model comparison (by AIC score, Fig. 3b) was the RLWM model with six free parameters: the RL learning rate α, WM capacity K, WM decay φ, WM prior weight ρ, positive learning bias parameter *bias,* and undirected noise ε. The inverse temperature parameter β was fixed to 100, as fitting a model with a free β, α, and ε led to worse precision in identifying all three parameters, and once said model was fit, only 1 out of 100 children, 2 out of 89 teens, and 0 out of 54 adults were best fit by that model in comparison to the other candidate models. Previous experience shows that fixing β in this model allows for best characterization of WM and RL parameters (Collins, 2018). Additionally, after running the same analyses with the model including a free β, all effects of age on other parameters remained, and there was no effect of age on β ($\rho = 0.07$, $p = 0.33$).

Instead, we chose to capture decision noise only with the ε undirected noise parameter. Because ε is a joint parameter applied to both the RL and WM processes, it is independent of learning and better recoverable. However, we do not claim that noise is of either form. Supplemental Figures 12 and 13 illustrate the final model recoverability and identifiability.

3.7. Model fitting procedure

We used the Matlab constrained optimization function fmincon to fit parameters (the Mathworks Inc., Natick, Massachusetts, USA). This was iterated with 20 randomly chosen starting points, to increase the likelihood of finding a global rather than local optimum. All parameters were fit with constraints [0 1], except the capacity parameter K. Due to the non-continuous nature of K, each set of random starting points was paired with each of the possible fixed values [2 3 4 5] of K. The best fit within those possible values of K was selected as a proxy for optimizing K alongside the other parameters.

3.8. Model comparison

We used the Akaike Information Criterion (AIC; Burnham & Anderson, 2002) to assess relative model fits and penalize model complexity. We previously showed that in the case of the RLWM model and its variants, AIC is a better approximation of model fit than Bayesian Information Criterion (BIC; Schwarz, 1978) at recovering the true model from generative simulations (Collins & Frank, 2012). Comparing RLWM and each of the variants of the simple RL model showed that RLWM provided a better fit to the data despite its additional complexity.

3.9. Model simulation

Model comparison alone is insufficient to assess whether the best fitting model sufficiently captures the data, as it provides only a relative measure of model fit (Nassar & Frank, 2016; Palminteri et al., 2017; Wilson and Collins, 2019, submitted). To test whether our models capture the key aspects of the behavior (i.e. learning curves), we simulated each model with fit parameters from each subject, with 100 repetitions per subject averaged to represent each subject's contribution to group-level behavioral effects (Fig. 3C, Supplemental Figure 11).

4. Results

4.1. Overall accuracy

To analyze coarse age effects on behavioral and modeling measures, we first grouped participants into three groups by age (8 to 12-year-olds, 13 to 17-year-olds, and 25 to 30-year-olds), and tested the continuous effect of age on performance for non-adult participants.

All participants performed significantly better than chance (33%, Fig. 2b). The mean accuracy for all subjects was 77.91% (median 80%). An ANOVA by age group revealed a main effect of group ($F(240) = 20.29$, $p = 9.10e^{-9}$). Post-hoc t-tests show that this main effect was driven by the differences in performance of the 8-12 year-old group from the 13-17 year-old group ($t(187) = 5.2$, $p < 10e^{-4}$) and the 25-30 year-old group ($t(152) = 5.2$, $p < 10e^{-4}$). 8 to 12-year-olds' performance was significantly worse overall (73% accuracy), while 13 to 17-year-olds and 25 to 30-year-olds performed similarly well (at 80.1% and 82.4% accuracy, respectively; $t(141) = 1$, $p = 0.3$). Within subjects aged 8 to 17, there was a positive correlation of age and overall accuracy (Fig. 2b; Spearman $\rho = 0.44$, $p = 2.e\text{-}10$), confirming an expected improvement in learning performance with age.

4.2. Learning

All age groups showed a similar qualitative pattern of learning whereby learning was faster in lower set sizes, a characteristic of combined working memory and reinforcement contributions to learning (Fig. 2a). Learning reached asymptote for set sizes 2 and 3 within the first three or four trials, and reached asymptote incrementally in set sizes 4 and 5. There was a strong negative effect of set size on performance (set size slope; see methods; $t(242) = 14.7$, $p = 2e\text{-}35$;207 of 243 participants). An ANOVA on the effect of set size revealed a main effect of age group ($F(240) = 7.6$, $p = 0.0006$). Post-hoc t-tests confirmed that this was driven by a larger negative effect of set size in 8 to 12-year-olds' performance than in 13 to 17-year-olds' ($t(187) = 2.85$, $p = 0.0049$) and adults' ($t(152) = 3.59$, $p = 0.0004$). 13 to 17-year-olds were not more or less affected by set size than 25 to 30-year-olds ($t(141) = 0.98$, $p = 0.33$). Within subjects aged 8 to 17, there was a negative correlation between age and set size slope ($\rho = -0.28$, $p = 8e\text{-}5$; Fig. 2b), supporting the previous analysis that participants' learning became less sensitive to set size with increased age.

4.3. Reaction time

Overall, mean reaction time (RT) decreased as a function of age ($\rho = -0.34$, $p = 2e\text{-}6$). RT variance also decreased as a function of age ($\rho = -0.27$, $p = 0.0002$). Reaction times were slower on high set size trials ($t(242) = 33.1$, $p = 0$; 242 out of 243 participants). There was no main effect of age group on the set size RT effect ($F(240) = 1.24$, $p = 0.29$).

All age groups and subjects appeared to be sensitive to the set size manipulation in both accuracy and reaction time, supporting the fact that all ages used both RL and WM to learn in this protocol. The 8- to 12-year-old group was slower, less accurate, and more sensitive to the set-size manipulation than participants in the group aged 13-17. Performance in the 13-17 year old group was comparable to adults. We next sought to use statistical and computational models to better characterize the underlying processes that drove the developmental changes in behavior.

4.4. Logistic regression

For each subject, we ran a trial-by-trial logistic regression predicting response accuracy with predictors set size, delay since the current stimulus was last presented (two potential markers of WM), previous correct trials for that stimulus, and previous incorrect trials (two potential markers of RL; see methods). Set size had a negative effect on performance in all age groups (8 to 12-year-olds $t = -6.96$, $p < 0.0003$; 13 to 17-year-olds $t = -4.39$, $p < 0.0003$; 25 to 30-year-olds $t = -3.88$, $p < 0.0003$). Delay also had a negative effect on performance in all groups (8 to 12-year-olds $t = -4.56$, $p < 0.0001$; 13 to 17-year-olds $t = -4.39$, $p < 0.0001$; 25 to 30-year-olds $t = -2.3$, $p = 0.026$). Number of previous correct trials had a positive effect on performance in all age groups (8 to 12-year-olds $t = 21.3$, $p < 0.0001$; 13 to 17-year-olds $t = 8.5$, $p < 0.0001$; 25 to 30-year-olds $t = 8.5$, $p < 0.0001$). Number of

previous incorrect trials had a negative effect in both 8-12 and 13-17 year olds but not in adults (8 to 12-year-olds $t = -11.1$, $p < 0.0001$; 13 to 17-year-olds $t = -4.93$, $p < 0.0001$; 25 to 30-year-olds $t = -1.1$, $p = 0.27$).

We tested the relationship between age and the individual logistic regression weights in a multiple regression predicting age from each individual's five regression weights. After excluding subjects with weights 2 standard deviations above or below the mean (10 8 to 17-year-olds), we found that the weight of the fixed effect ($t(173) = 7.1$, $p < 10e\text{-}4$), and the weight of previous correct trials ($t(173) = -2.9$, $p = 0.004$) predicted age. None of the other predictors (pinc, delay, ns) predicted age (p's > 0.09).

Results so far confirmed our prediction that younger participants would perform worse than older participants. There was evidence for successful RL recruitment in all subjects based on use of previous correct feedback, as well as evidence of WM recruitment based on set-size and delay effects. However, results so far were ambiguous as to whether WM, RL, or both drove learning improvement with age. While a decrease in set size effect could hint at a WM effect, it is equally possible that worse performance in high set sizes in younger children could be due to worse RL (as hinted by the logistic regression results), which can support learning when WM is unavailable (e.g. at high set sizes). To clarify these findings, we next sought to model individual participants' behavior with a mixture model capturing both RL and WM contributions to learning (see Methods).

4.5. Modeling

We fit participant behavior with a two-module reinforcement learning and working memory model. Both modules update the value of stimulus-action pairings and contribute to choice. While RL is slow to learn, the WM module learns quickly and is also subject to forgetting or interference. Furthermore, our model assumes that the WM module's contribution to choice diminishes with load, in accordance to its limited capacity. Thus, the WM module captures quick learning early on in a block, especially in low set sizes, while the RL module can account for slower learning leading to stable asymptotic performance, especially in high set size blocks. See Methods for model details. Model comparison favored the RLWM model over other candidate models in all age groups (Fig. 3ab; see methods). The exceedance probability in favor of the RLWM model was 1 in all groups (Rosa et al., 2010). RLWM was also the best model out of 6 candidate models for 83 out of 100 8 to 12-year-olds, 58 out of 59 13 to 17-year-olds, and 25 out of 54 25 to 30-year-olds (Fig. 3a). Model simulations with fit parameters reproduced subject behavior, as well as differences between age groups (Fig. 3C, Supplemental Fig. 1, Supplemental Fig. 11).

We first investigated two noise parameters (Fig. 4: Joint, epsilon and bias) and as expected, there was an effect of age group on the decision noise parameter epsilon (Kruskall-Wallis $p = 0.025$). Post-hoc comparison revealed that this was driven by the separation in behavior between children 8-12 and the older age groups (Rank-sum test 8 to 12-year-olds vs. 13 to 17-year-olds: $p = 0.0003$; 8 to 12-year-olds vs. 25 to 30-year-olds: $p = 0.002$; 13 to 17-year-olds vs. 25 to 30-year-olds: $p = 0.94$). There was a negative relationship between age and decision noise in the 8 to 17-year-old sample ($\rho = -0.25$, $p = 0.0005$), in boys ($\rho = -0.22$, $p = 0.024$), and in girls ($\rho = -0.3$, $p = 0.005$), showing that decisions were less noisy in older participants. There was also an effect of age group on the bias parameter (Kruskal-Wallis $p = 0.0003$). Post-hoc comparisons also revealed that the difference between 8 to 12-year-olds and the other age groups drove the main effect of group (Rank-sum test 8 to 12-year-olds vs. 13 to 17-year-olds: $p < 10e\text{-}4$; 8 to 12-year-olds vs. 25 to 30-year-olds: $p = 0.001$; 13 to 17-year-olds vs. 25 to 30-year-olds: $p = 0.67$). In the 8 to 17-year-old sample, there was also a negative relationship between age and bias ($\rho = -0.3$, $p = 1.9e\text{-}07$). This relationship was present separately in both boys ($\rho = -0.3$, $p = 0.002$) and girls ($\rho = -0.44$, $p = 1.9e\text{-}05$),

We next investigated the relationship between age and RL learning rate α. There was a robust effect of age group on the RL learning rate parameter (Fig. 4: RL; Kruskal-Wallis $p = 0.0002$). Additional comparison between groups showed that 8 to 12-year-olds had a lower average learning rate than 13 to 17-year-olds ($p < 10e\text{-}4$) and 25 to 30-year-olds ($p < 10e\text{-}4$). 13 to 17-year-olds did not differ from 25 to 30-year-olds ($p = 0.3$). Individual learning rates showed a significant upward trend with age 8-17 ($\rho = 0.31$, $p = 2e\text{-}05$), which was significant separately in both boys ($\rho = 0.32$, $p = 0.001$) and girls ($\rho = 0.29$, $p = 0.007$).

Finally, we investigated WM parameters. Surprisingly, we found no significant differences across our coarse age groups in WM capacity (Kruskal-Wallis test, $p = 0.83$), (Fig. 4: WM left) (although see PDS and

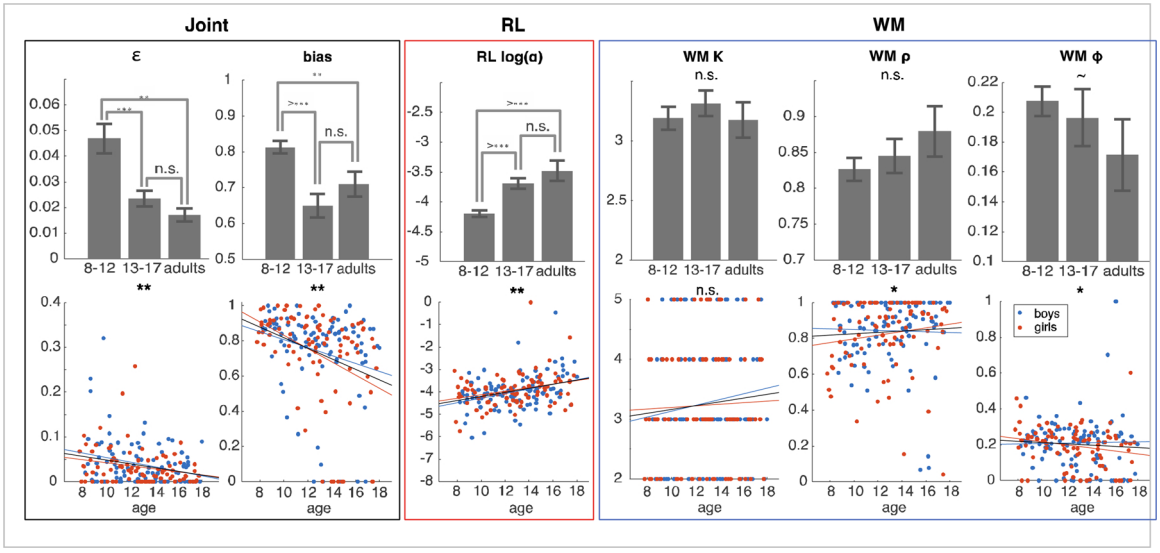


Fig. 4. Effects of age on RLWM model parameters. There was an effect of both age group and age in years on the bias and ε decision noise parameters, whereby 8-12-year-olds had noisier behavior, and integrated negative feedback less than 13-17-year-olds and 25-30-year-olds. There were robust effects of age on the RL learning rate parameter α. There was no effect of coarse age groups or continuous effect of age within the 8-17 sample on the WM capacity parameter, and weak effects on WM weight ρ and decay φ. Error bars on bar graphs are standard error of the mean for each age group. ~ indicates marginal significance at the $p < 0.1$ level, * indicates $p < 0.05$, ** indicates $p < 0.01$, n.s. stands for not significant.

testosterone group results below). When we focused on individual subjects aged 8 to 17, we found no monotonic relationship between WM capacity and age (Fig. 4: WM; Spearman $\rho = 0.09$, $p = 0.23$). We found no differences across coarse age groups in WM weight (Fig. 4: WM; $p = 0.25$), but a marginal effect on WM decay (Fig. 4; $p = 0.08$). These effects were stronger when investigating the continuous effect of age within non-adults: There was a small positive relationship between age and WM weight (Spearman $\rho = 0.15$, $p = 0.04$; Boys: $\rho(101) = 0.06$, $p = 0.58$; Girls: $\rho(86) = 0.25$, $p = 0.02$). There was also a negative relationship between age and WM decay (Spearman $\rho = -0.17$, $p = 0.02$) that was inconsistent across genders (Boys: $\rho = -0.08$, $p = 0.42$; Girls: $\rho = -0.28$, $p = 0.01$). This continuous effect was mostly driven by the youngest participants, with more WM decay and less WM weight.

Children of the same age can differ in their stage of pubertal maturation with considerable individual variability. There are also notable sex differences in the timing of pubertal onset. Pubertal onset and later pubertal milestones may also produce non-monotonic changes over time such as a step or 'inverted U' shaped curve (Braams et al., 2015; Piekarski, Boivin, et al., 2017; Piekarski et al., 2017b). Gonadal hormones, such as testosterone and its metabolites may also be playing an activational role in cortical or basal ganglia function on the day of testing (Delevich, Piekarski, et al., 2019). Therefore, to explore developmental changes in learning with a finer resolution relevant to puberty we divided the 8-17 year old sample into 4 evenly divided bins first by age (<10.5, 10.6-12.8, 12.9-14.8, and >14.9 years old) then by pubertal development scale (PDS) (roughly pre-, early, mid-, and late/post-pubertal) and then by salivary testosterone (low, low-mid, mid-high, and high levels within gender) at time of test (T1) (see methods, Supplemental material). Cross-sectional data are quite limiting in differentiating age and pubertal effects, however we carefully explored the data and the effects of age, PDS, and testosterone concentration.

Division by finely graded age bins showed that noise and bias decreased with age (Supplementary Figures S4, S5; Noise: chi-squared = 15.69, $p = 0.0013$; bias: chi-squared = 24.1, $p = 2e\text{-}05$) and RL learning rate increased significantly with age (Supplementary Figure S6; chi-squared = 19.3, $p = 0.0002$) with no inverted U patterns. This general pattern was present in both boys (Supplementary Figures S4, S5, and S6; RL learning rate: chi-squared = 12.1, $p = 0.007$; noise: chi-squared = 6.7, $p = 0.08$; bias: chi-squared = 9.1, $p = 0.028$) and girls (RL learning rate: chi-squared = 8.1, $p = 0.044$; noise: chi-squared = 10.13, $p = 0.018$; bias: chi-squared = 26.2, $p = 0.001$). We observed similar results when non-adult participants were divided into bins based on PDS or testosterone measures (see supplementary table

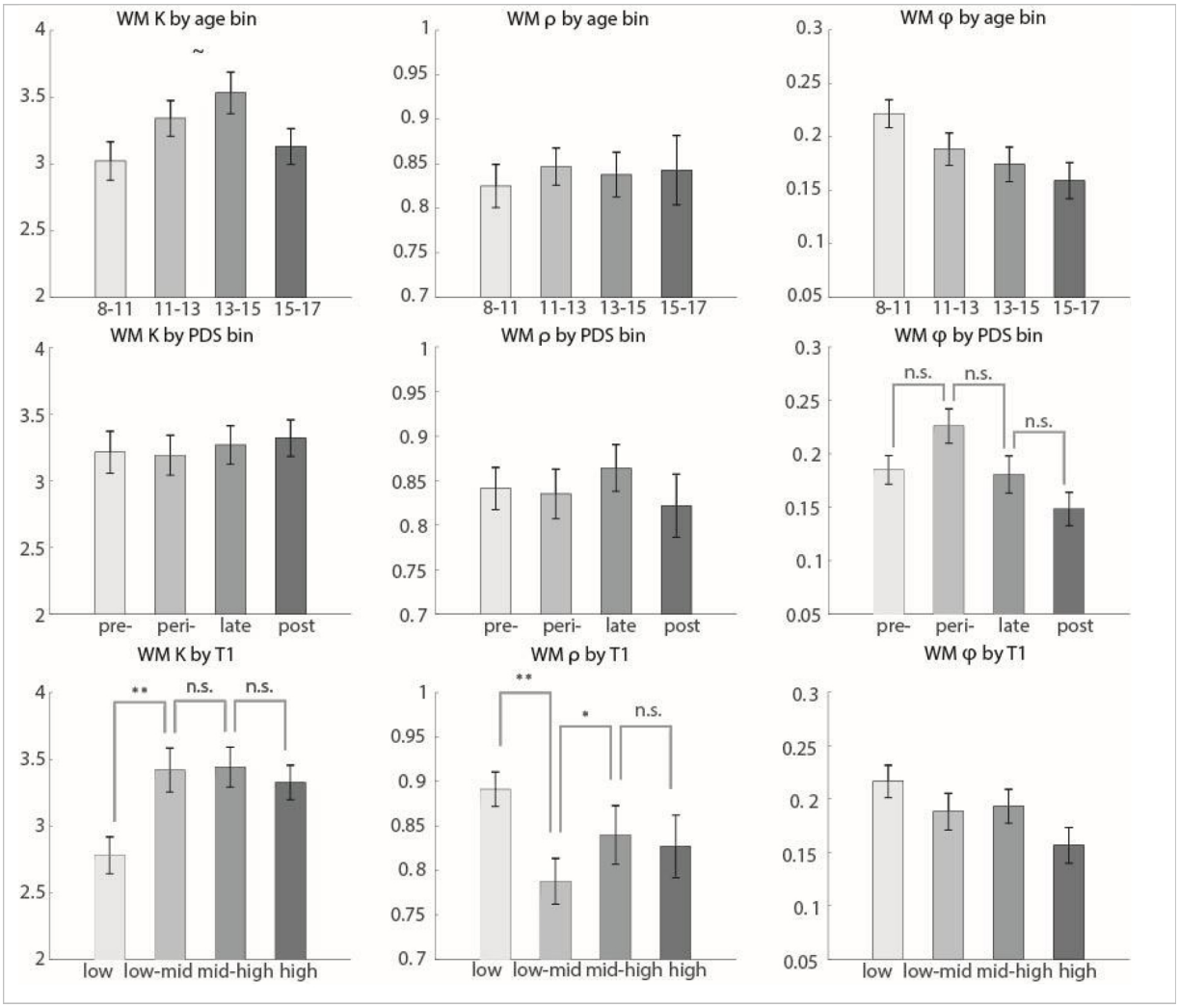


Fig. 5. WM parameters by age, PDS, and sample 1 testosterone bins. All subjects aged 8 to 17 were binned according to age, pubertal development score (PDS), and salivary testosterone from the in-lab sample (T1). Girls and boys were binned separately according to gender-specific quartiles, then combined into equal-sized groups for each measure. Effects of group were assessed using the non-parametric Kruskal-Wallis test. Tests for which the group effect was significant at $p < 0.05$ were further examined with post-hoc non-parametric t-tests. n.s. indicates not significant, ~indicates marginal significance at $p < 0.1$ level, * $p < 0.05$, ** $p < 0.01$. Bin means are plotted; error bars show the standard error of the mean for each bin.

S2). Age, PDS and mean salivary testosterone were all also highly correlated (Supplementary Figure S2), therefore it was not clear if puberty onset played a significant role in RL-related changes or not.

We next examined the relationship between finely graded groups based on age, PDS and salivary testosterone measures and WM parameters. Here the greater number of groups enabled observation of 'U' or 'inverted U shapes.' Patterns of this type were apparent when subjects were grouped by PDS or T1. We had predicted potential effects of puberty onset on WM due to our previous work on effects of gonadal hormones on inhibitory neurotransmission in the PFC of rodent models (Piekarski, Boivin, et al., 2017; Piekarski, et al., 2017b).

Subjects that were grouped by PDS did show significant differences in WM decay that was stronger in boys than girls (see Fig. 5, Supplemental Figure 8; all: chi-squared = 10.02, p = 0.018; boys: chi-squared = 15.09, p = 0.0017; girls: chi-squared = 5.51, p = 0.14). Changes were strongest at puberty onset. We also uncovered a significant effect of testosterone at time of test (T1) on WM capacity (see Fig. 5, Supplemental Figure 9; chi-squared = 13.15, p = 0.0043). This effect was marginal in boys alone (chi-squared = 6.61, p = 0.085) and significant in girls alone (chi-squared = 7.86, p = 0.049). When subjects were grouped by age (in 4 bins age 8-17), without regard to PDS or T, no significant group differences in WM parameters were found (though there was a marginal inverse U-shaped effect on WM capacity, and decreasing effect on decay; Fig. 5, Supplemental Figures 8 and 9).

These results support our first prediction that changes in both RL and WM processes separately drive learning changes in subjects aged 8 to 30. However, we were surprised to find that changes in RL drive changes in learning over a longer period of development than changes in WM, in opposition to our second prediction. Finally, we support our third prediction that WM processes would differ in groups separated by pubertal development and testosterone.

5. Discussion

In this study, we examined developmental changes in the working memory (WM) and reinforcement learning (RL) processes that contribute to simple stimulus-action association learning. While many developmental models emphasize late prefrontal and parietal cortex maturation and gains in WM in late adolescence (Huizinga et al., 2006; Larsen & Luna, 2018), previous studies have also shown developmental changes in reinforcement learning parameters between mid-adolescence and adulthood (Palminteri et al., 2016; Potter et al., 2017; Van Den Bos et al., 2012). However, these studies were not able to disentangle WM and RL contributions to learning. Our task and computational model were designed to address how WM and RL jointly contribute to learning in a simple task (Collins, 2018; Collins et al., 2017; Collins & Frank, 2012; Collins et al., 2017). The task and model are also notable because they reveal that WM is recruited in even simple reward learning tasks that are often assumed to include only model-free RL processes.

Here we used the RLWM task and computational methods to measure development of RL- and WM-based processes in youth aged 8 to 17 and adults aged 25 to 30. We found that when we grouped subjects into 8 to 12-year-olds, 13 to 17-year-olds, and 25 to 30-year-olds all recruited both RL and WM in parallel for learning. 13 to 17-year-olds' performance closely approximated 25 to 30-year-olds', while participants aged 8 to 12 learned more slowly and reached a lower asymptotic accuracy. Using our standard analyses and computational models, we next sought to disentangle changes in RL and WM systems' contributions to gains in learning and also isolate contributions of decision noise and learning biases.

6. RL and WM development

We found the RL learning rate steadily increased throughout adolescence (Fig. 4) enabling youth to reach adult-like levels of performance in the 13 to 17-year-old group. Given that all correct choices were rewarded 100% of the time and there was an unchanging reward structure, an optimal strategy would be to learn quickly from rewards (i.e. have a higher learning rate). This allows learners to reach asymptotic accuracy faster as long as the environment is stable. The developmental change in learning rate allowed the 13 to 17-year-old group and 25 to 30-year-old group to learn stimulus-action associations faster than the 8 to 12-year-old group, which benefitted their overall task performance.

In our three WM parameters, changes were more modest. We did observe small effects of age on WM weight and WM decay in 8 to 17-year-olds (Fig. 4), reflecting increased use and stability of WM with increasing age. Changes in WM capacity were only observed at puberty onset in early adolescence (Fig. 5) when youth were grouped by PDS or testosterone concentration quartiles (Fig. 5).

Indeed, our behavioral results and computational model did not support our hypotheses that WM would show protracted maturation and serve as the main driver of learning improvement from 8 to 17 years of age. Instead we found that RL-based learning showed protracted development until age 17. Unpacking what this means for performance, in the raw data we found that 8 to 12-year-olds' behavior was more sensitive to the set size or learning load. We attribute this difference to a weaker RL system that was unable to make up for WM's limitations in high set size conditions, more than a weaker capacity-limited WM.

7. Noise and bias

Our data also isolated noise and a positive learning bias as important variables in the development of simple associative learning ages 8 to 18. Across each of our analyses, decision noise consistently decreased with age. 8- to 12-year-olds' behavior was noisier than older subjects', as evidenced by their failure to reach the same asymptotic accuracy as 13- to 30-year-olds, the relatively small intercept term from the logistic regression on accuracy, and the decrease in the epsilon decision noise parameter with age in years. Each of these results emphasizes the 8- to 12-year-olds' tendency towards attentional lapse and higher probability of disregarding learned information to choose randomly. It may also be indicative of a higher propensity for exploration (Schulz et al., 2019, bioRxiv preprint.; Somerville et al., 2017). 13- to 17-year-olds and 25- to 30-year-olds exhibited lesser amounts of decision noise and tended more towards making decisions based on learned value information. This finding is consistent with other work showing a decrease in both neural and behavioral noise in learning and memory processes in adolescence (Montez et al., 2017, 2019; Somerville et al., 2017). We also found that there was a negative relationship of age and a bias towards learning from positive feedback, which is consistent with other work on the development of reinforcement learning (Davidow et al., 2016; Hauser et al., 2015; Jones et al., 2014; Palminteri et al., 2016; Van Den Bos et al., 2012). Our task and model are not designed to disentangle more subtle aspects of noise and learning bias, but our data indicate that they too contribute to changes in learning and are therefore are important to consider in the developmental science of learning.

8. Limitations of our data

One limitation of our findings was a difference in socioeconomic status (SES) between our adult group ages 25 to 30 and our non-adult group ages 8 to 17. SES was measured by the proxy of self-reported household income. Excluding the subjects who didn't disclose their socioeconomic status (n = 88), a kruskal-wallis test revealed an effect of group (chi-squared = 56.15, p = 6.4e-13; Supplementary Figure 10). Post-hoc t-tests across groups revealed that adults were at a different income level than both 8 to 12-year-olds (p = 1.5e-12) and 13 to 17-year-olds (p = 6.0e-18). 8 to 12-year-olds and 13 to 17-year-olds were at the same income level (p = 0.15). This could limit the interpretation of our results showing no difference between 13 to 17-year-olds and adults. Indeed, children of lower socioeconomic status have been shown to score lower on tests of executive function and working memory than

their peers (Farah et al., 2006; Hackman & Farah, 2009; Noble et al., 2007; Noble et al., 2005) and these effects can carry over into adulthood (Evans & Schamberg, 2009; Hackman, Farah, & Meaney, 2010; Oshri et al., 2019). However, this confound may be mitigated by the limitations of using household income as a proxy for socio-economic status. Although reporting low incomes, the young adult subjects also showed high levels of education – all adult subjects had completed high school and 50 out of 54 had some college or higher education. This might indicate that our adult participants might be in early career stages, and that their current income does not reflect the socio-economic status we might assign to their past or future household income. There were no household income differences between the 8 to 12-year-olds and 13 to 17-year-olds, where we found most differences in task performance.

Our study used a cross-sectional design and measures of puberty at only one timepoint in a complex multi-component process. Pubertal development was highly correlated with age in our sample, however there were late and early developing individuals within our sample. Pubertal tempo, the speed at which children complete stages of puberty, was not captured by our data. Thus, findings from our study, particularly those that implicate puberty onset and hormone levels as potential mechanisms should be followed up with longitudinal studies.

9. Interpretation of our data

Late adolescent maturation is widely associated with the development of higher cognitive resources, including adult-like working memory (Crone et al., 2006; Geier et al., 2009; Huizinga et al., 2006; Luna, 2009; McAuley & White, 2011) and the use of increasingly complex learning strategies (Crone et al., 2006; Decker et al., 2016; Potter et al., 2017; Selmeczy et al., 2018). The development of these forms of higher cognition are most often attributed to the development of prefrontal cortex, which does not reach functional or structural maturation until late in the second decade or even middle of the third decade of life (Bunge et al., 2002; Casey et al., 2008). Therefore, it was somewhat surprising that we detected no age-related changes in WM capacity in subjects aged 8 to 17 despite our large sample. This is partially mitigated by PDS and testosterone-related findings that show changes in WM capacity with the transition into puberty (Fig. 5) and the fact that other aspects of WM (decay and weight) did show significant if weak effects of age (Fig. 4 & 5).

To better integrate this work into the established literature on the developmental science of learning, it may be helpful to separate these three different aspects of WM and to dissociate 1) the use of WM to maintain information from 2) the use of WM to manipulate information. It is possible the use of WM to manipulate information may develop in a more extended fashion over adolescence (Crone et al., 2006), while the use of WM to maintain information perhaps develops earlier. Recent results with a simple WM assay, a verbal span task, found age groups aged 9 to 25 performed equally well (Potter et al., 2017) and thus found no developmental effects on WM. In our RLWM task WM is used to maintain stimulus-action associations, however, it is debatable the extent to which the subjects need to simply maintain or manipulate information. In our task, information was not explicitly given to participants for them to hold, but they needed to integrate three temporally separate pieces of information (the stimulus observed, the choice made, and the feedback received) to determine the relevant information to maintain in WM. Thus, we would argue RLWM WM still constitutes a sophisticated use of WM.

While dual systems models of development (Shulman et al., 2016) have often highlighted that subcortical areas mature before prefrontal cortex, there is also an important literature showing changes in striatal function through adolescence. Adolescents in the mid-teen years have been shown to be more sensitive to and motivated by rewards than pre-pubertal children and post-pubertal adults (Casey et al., 2008; Braams et al., 2015; Davidow et al., 2016; Gulley et al., 2018; Somerville et al., 2010). This increased motivation is reflected both in BOLD activation of the reward-responsive ventral striatum (Braams et al., 2015; Galvan et al., 2006b) and in increased recruitment of areas of the frontoparietal cognitive control network, when necessary (Somerville & Casey, 2010). This could partially account for our results, explaining why 13 to 17-year-olds use RL more efficiently than 8 to 12-year-olds. However, such a relationship would predict differences between 13 to 17-year-olds and 25 to 30-year-olds, as seen in learning tasks that manipulate reinforcement more directly (Palminteri et al., 2016). We did not observe any difference between subjects aged 13 to 17 and adults in both behavior and in modeling; this might be due to a comparatively weak reinforcement manipulation, though it was sufficient to reveal strong differences between 8 to 12-year-olds and 13 to 17-year-olds.

The lack of a consistent relationship of age and WM capacity in this sample could also be interpreted as reflecting the intrinsic difficulty of perfectly disentangling the RL and WM systems. Indeed, treating the BG-dependent RL system and PFC-parietal dependent WM system as distinct, non-overlapping, independent systems is overly simplistic (Casey et al., 2016; Shulman et al., 2016). There is a large literature highlighting the role of striatal dopamine in working memory gating in PFC (Cools, 2011; Frank et al., 2001; O'Reilly & Frank, 2014; Gruber et al., 2006; Hazy et al., 2007). Furthermore, recent work from our and others' labs shows potential interactions between these systems that go further than competition for choice (Collins, 2018; Collins et al., 2017; Daw et al., 2011; Gold et al., 2017; Sharpe et al., 2017). Nevertheless, there is also ample evidence that these systems are, on first order, separable with differing contributions to learning. Contrasting the strong statistical effects we observed for developmental trends in RL, bias and noise parameters, to the null or weak results obtained for WM parameters weakens other possible interpretations of our findings in terms of overlap: any shared subcortical effects driving RL changes should then also drive WM changes. Thus, we believe that the absence of strong effects of age on WM parameters as going against our original prediction, and showing that developmental changes in RL are more influential on changes in learning during adolescence.

It is also important to consider the role episodic memory might be playing in our task. Recent research showed that a component of learning from reinforcement can be accounted for by episodic memory sampling (Bornstein & Daw, 2012; Bornstein & Norman, 2017; Myers et al., 2002). In fact, previous versions of this task found hippocampal contributions to learning (Collins et al., 2017). Thus, it is possible that some behavior captured by the RL component of the model actually included contributions of the episodic memory system, not just the reward learning systems. Some facets of episodic memory, like recognition memory, have been shown to develop by age 8, much earlier than prefrontal cortex-dependent executive function (Ghetti & Bunge, 2012). However, recent work shows that the interactions between hippocampus and prefrontal cortex – crucial for efficient memory search and retrieval – continue developing into adolescence (Murty et al., 2016; Selmeczy et al., 2018). Further, the development of both systems allows for the integration of relevant past experiences with goal-directed attention and can influence choice in the reward-driven BG (Murty et al., 2016). This developmental time course might account for some of the changes attributed to RL learning rate in our results. It will be an important topic for future research to better characterize episodic memory contributions to learning, in parallel to WM and RL contributions.

10. Conclusions

Using a sample of 241 participants 8-17 and 25-30 years old, we aimed to characterize the distinct developmental trajectories of two cognitive systems that contribute to learning in parallel, even in very simple situations: the RL system and WM system. Performance in a simple stimulus-action learning task which manipulated cognitive load revealed broad differences between subjects aged 8 to 12 and subjects 13 to 17, whose raw performance was adult-like. While 8 to 12-year-olds

were noisier in all conditions, their learning suffered more from a load increase than 13 to 17-year-olds' and 25 to 30-year-olds'. In all participants, RL learning compensated for WM limitations under higher load. Computational modeling revealed that the stronger effect of load in 8 to 12-year-olds was best explained by weaker RL compensation. There was evidence of subtle gains in WM development 8-17 in terms of WM use and decay, but the effect size was weak. Changes in WM capacity were only apparent at puberty onset when subjects were sorted by PDS or testosterone levels. These results were surprising based on the established literature on the late anatomical development of the prefrontal and parietal association cortices. This work highlights the importance of carefully accounting for multiple systems' contributions to learning when assessing group and individual differences and suggests that the development of reinforcement learning processes plays a protracted role in changes in learning during adolescent development. We hope these findings can fruitfully inform educational methods and intervention work. Future research in the science of learning should aim to develop experimental paradigms and computational models that more precisely define and dissociate different sources of noise and the use of reinforcement learning, working memory, and episodic memory throughout development.

Funding

This work was funded by National Science Foundation SL-CN grant #1640885 to RD, AGEC, and LW.

The authors declare no conflict of interest.

Acknowledgements

We gratefully acknowledge the numerous people who contributed to this research: Lance Kriegsfeld, Celia Ford, Jennifer Pfeifer, Megan M. Johnson, Ahna Suleiman, Silvia Bunge, Liyu Xia, Rachel Arsenault, Josephine Christon, Shoshana Edelman, Lucy Eletel, Haley Keglovits, Julie Liu, Justin Morillo, Nithya Rajakumar, Nick Spence, Tanya Smith, Benjamin Tang, Talia Welte, Lucy B. Whitmore, and Amy Zou. We are grateful to our participants and their families for their time and participation in this study.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.dcn.2019.100732.

References

Badre, D., Kayser, A.S., D'Esposito, M., 2010. Frontal cortex and the discovery of abstract action rules. Neuron 66 (2), 315–326. https://doi.org/10.1016/j.neuron.2010.03.025.

Bornstein, A.M., Daw, N.D., 2012. Dissociating hippocampal and striatal contributions to sequential prediction learning. European Journal of Neuroscience 35 (7), 1011–1023. https://doi.org/10.1111/j.1460-9568.2011.07920.x.

Bornstein, A.M., Norman, K.A., 2017. Reinstated episodic context guides sampling-based decisions for reward. Nature Neuroscience 20 (7), 997–1003. https://doi.org/10.1038/nn.4573.

Braams, B.R., van Duijvenvoorde, A.C.K., Peper, J.S., Crone, E.A., 2015. Longitudinal Changes in Adolescent Risk-Taking: A Comprehensive Study of Neural Responses to Rewards, Pubertal Development, and Risk-Taking Behavior. Journal of Neuroscience 35 (18), 7226–7238. https://doi.org/10.1523/jneurosci.4764-14.2015.

Bunge,, A., Dudukovic,, N.M., Thomason,, M.E., Vaidya,, C.J., Gabrieli,, J.D.E., 2002. Immature Frontal Lobe Contributions to Cognitive Control in Children: Evidence from fMRI. Neuron 33, 301–311. https://doi.org/10.1016/S0896-6273(01)00583-9.

Casey, B.J., Galván, A., Somerville, L.H., 2016. Beyond simple models of adolescence to an integrated circuit-based account: A commentary. Developmental Cognitive Neuroscience 17, 128–130. https://doi.org/10.1016/j.dcn.2015.12.006.

Casey, B.J., Jones, R.M., Hare, T.A., 2008. The adolescent brain. Annals of the New York Academy of Sciences 1124, 111–126. https://doi.org/10.1196/annals.1440.010.

Cohen, J.R., Asarnow, R.F., Sabb, F.W., Bilder, R.M., Bookheimer, S.Y., Knowlton, B.J., Poldrack, R.A., 2010. A unique adolescent response to reward prediction errors. Nature Neuroscience 13 (6), 669–671. https://doi.org/10.1038/nn.2558.

Collins, A.G.E., 2018. The tortoise and the hare: Interactions between reinforcement learning and working memory. Journal of Cognitive Neuroscience 30, 1422–1432.

Collins, A.G.E., Ciullo, B., Frank, M.J., Badre, D., 2017. Working Memory Load Strengthens Reward Prediction Errors. The Journal of Neuroscience 37 (16), 4332–4342. https://doi.org/10.1523/jneurosci.2700-16.2017.

Collins, A.G.E., Frank, M.J., 2012. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. European Journal of Neuroscience 35 (7), 1024–1035. https://doi.org/10.1111/j.1460-9568.2011.07980.x.

Collins, A.G.E., Frank, M.J., 2014. NIH Public Access 120 (1), 190–229. https://doi.org/10.1037/a0030852.Cognitive.

Collins, A.G.E., Frank, M.J., 2018. Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.1720963115, 201720963.

Cools, R., 2011. Dopaminergic control of the striatum for high-level cognition. Current Opinion in Neurobiology 21 (3), 402–407. https://doi.org/10.1016/J.CONB.2011.04.002.

Crone, E.A., Wendelken, C., van Leijenhorst, L., Donohue, S., Bunge, S.A., 2006. Neurocognitive development of the ability to manipulate information in working memory. Proceedings of the National Academy of Sciences 103 (24), 9315–9320. https://doi.org/10.1073/pnas.0510088103.

Curtis, C.E., D'Esposito, M., 2003. Persistent activity in the prefrontal cortex during working memory. Trends in Cognitive Sciences 7 (9), 415–423. https://doi.org/10.1016/S1364-6613(03)00197-9.

D'Esposito, M., Postle, B.R., 2015. The Cognitive Neuroscience of Working Memory. Annu Rev Psychol. 66 (66), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031.THE.

Davidow, J.Y., Foerde, K., Galván, A., Shohamy, D., 2016. An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. Neuron 92 (1), 93–99. https://doi.org/10.1016/j.neuron.2016.08.031.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J., 2011. Model-based influences on humans' choices and striatal prediction errors. Neuron 69 (6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027.

Decker, J.H., Otto, A.R., Daw, N.D., Hartley, C.A., 2016. From Creatures of Habit to Goal-Directed Learners. Psychological Science 27 (6), 848–858. https://doi.org/10.1177/0956797616639301.

Delevich, K., Piekarski, D., Wilbrecht, L., 2019. Neuroscience: Sex Hormones at Work in the Neocortex. Current Biology 29 (4), R122–R125. https://doi.org/10.1016/j.cub.2019.01.013.

Delevich, K., Thomas, A.W., Wilbrecht, L., 2019. Adolescence and "Late Blooming" Synapses of the Prefrontal Cortex. Cold Spring Harbor Symposia on Quantitative Biology, 037507. https://doi.org/10.1101/sqb.2018.83.037507. LXXXIII.

Evans, G.W., Schamberg, M.A., 2009. Childhood poverty, chronic stress, and adult working memory. Proceedings of the National Academy of Sciences of the United States of America 106 (16), 6545–6549. https://doi.org/10.1073/pnas.0811910106.

Farah, M.J., Shera, D.M., Savage, J.H., Betancourt, L., Giannetta, J.M., Brodsky, N.L., Hurt, H., 2006. Childhood poverty: Specific associations with neurocognitive development. Brain Research 1110 (1), 166–174. https://doi.org/10.1016/j.brainres.2006.06.072.

Frank, M.J., Loughry, B., Reilly, R.C.O., 2001. Interactions between frontal cortex and basal ganglia in working memory: A computational model. Cognitive, Affective, & Behavioral Neuroscience 1 (co(2)), 137–160.

Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T., Hutchison, K.E., 2007. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proceedings of the National Academy of Sciences of the United States of America 104 (41), 16311–16316. https://doi.org/10.1073/pnas.0706111104.

O'Reilly, R., Frank, M.J., 2014. Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. Neural Computation 1–26. Retrieved from https://www.mitpressjournals.org/doi/abs/10.1162/089976606775093909.

Galvan, A., Hare, T.A., Parra, C.E., Penn, J., Voss, H., Glover, G., Casey, B.J., 2006. Earlier Development of the Accumbens Relative to Orbitofrontal Cortex Might Underlie Risk-Taking Behavior in Adolescents. Journal of Neuroscience 26 (25), 6885–6892. https://doi.org/10.1523/JNEUROSCI.1062-06.2006.

Galvan, A., Hare, T.A., Parra, C.E., Penn, J., Voss, H., Glover, G., Casey, B.J., 2006. Earlier Development of the Accumbens Relative to Orbitofrontal Cortex Might Underlie Risk-Taking Behavior in Adolescents. Journal of Neuroscience 26 (25), 6885–6892. https://doi.org/10.1523/JNEUROSCI.1062-06.2006.

Galvan, Adriana, Delevich, K., Wilbrecht, L., 2019. Cortico-Striatal Circuits and Changes in Reward, Learning, and Decision Making in Adolescence. In: Gazzaniga (Ed.), The Cognitive Neurosciences, 6th edition.

Geier, C.F., Garver, K., Terwilliger, R., Luna, B., 2009. Development of Working Memory Maintenance. Journal of Neurophysiology 101 (1), 84–99.

Ghetti, S., Bunge, S.A., 2012. Neural changes underlying the development of episodic memory during middle childhood. Developmental Cognitive Neuroscience 2 (4), 381–395. https://doi.org/10.1016/j.dcn.2012.05.002.

Giedd, J.N., Blumenthal, J., Jeff ries, N.O., et al., 1999. Brain development during childhood and adolescence: a longitudinal MRI study. Nature Neurosci 1999. Nature Neuroscience 2, 861–863, 2, 861–863.

Gold, J.M., Waltz, J.A., Frank, M.J., Albrecht, M.A., Collins, A.G.E., 2017. Interactions Among Working Memory, Reinforcement Learning, and Effort in Value-Based Choice: A New Paradigm and Selective Deficits in Schizophrenia. Biological Psychiatry 82 (6), 431–439. https://doi.org/10.1016/j.biopsych.2017.05.017.

Gruber, A.J., Dayan, P., Gutkin, B.S., Solla, S.A., 2006. Dopamine modulation in the basal ganglia locks the gate to working memory. Journal of Computational Neuroscience 20 (2), 153–166. https://doi.org/10.1007/s10827-005-5705-x.

Gulley, J.M., Willing, J., Paul, M.J., Flores, C., Walker, D.M., Bell, M.R., 2018. Adolescence and Reward: Making Sense of Neural and Behavioral Changes Amid the Chaos. The Journal of Neuroscience 37 (45), 10855–10866. https://doi.org/10.1523/jneurosci.1834-17.2017.

Hackman, D.A., Farah, M.J., 2009. Socioeconomic status and the developing brain. Trends in Cognitive Sciences 13 (2), 65–73. https://doi.org/10.1016/j.tics.2008.11.003.

Hackman, D.A., Farah, M.J., Meaney, M.J., 2010. Socioeconomic status and the brain: mechanistic insights from human and animal research. Nature Reviews Neuroscience 434 (1957), 651–659. Retrieved from https://doi.org/10.1038/nrn2897.

Hauser, T.U., Iannaccone, R., Walitza, S., Brandeis, D., Brem, S., 2015. Cognitive flexibility in adolescence: Neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. NeuroImage 104, 347–354. https://doi.org/10.1016/J.NEUROIMAGE.2014.09.018.

Hazy, T.E., Frank, M.J., O'Reilly, R.C., 2007. Toward an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. Modelling Natural Action Selection 239–263. https://doi.org/10.1017/CBO9780511731525.016. April.

Herting, M.M., Gautam, P., Spielberg, J.M., Kan, E., Dahl, R.E., Sowell, E.R., 2014. The role of testosterone and estradiol in brain volume changes across adolescence: A longitudinal structural MRI study. Human Brain Mapping 35 (11), 5633–5645. https://doi.org/10.1002/hbm.22575.

Huizinga, M., Dolan, C.V., van der Molen, M.W., 2006. Age-related change in executive function: Developmental trends and a latent variable analysis. Neuropsychologia 44 (11), 2017–2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010.

Jones, R.M., Somerville, L.H., Li, J., Ruberry, E.J., Powers, A., Mehta, N., Casey, B.J., 2014. Adolescent-specific patterns of behavior and neural activity during social reinforcement learning. Cognitive Affective Behavioral Neuroscience 14 (2), 683–697. https://doi.org/10.3758/s13415-014-0257-z.

Juraska, J., Willing, J., 2017. Pubertal onset as a critical transition for neural development and cognition. Brain Research 1654 (B), 87–94. https://doi.org/10.1016/j.brainres.2016.04.012.

Katahira, K., 2015. The relation between reinforcement learning parameters and the influence of reinforcement history on choice behavior. Journal of Mathematical Psychology. https://doi.org/10.1016/j.jmp.2015.03.006.

Klingberg, T., 2006. Development of a superior frontal-intraparietal network for visuo-spatial working memory. Neuropsychologia 44 (11), 2171–2177. https://doi.org/10.1016/j.neuropsychologia.2005.11.019.

Klingberg, T., Forssberg, H., Westerberg, H., 2002. Increased Brain Activity in Frontal and Parietal Cortex Underlies the Development of Visuospatial Working Memory Capacity during Childhood. Journal of Cognitive Neuroscience 718, 1–10.

Larsen, B., Luna, B., 2018. Adolescence as a neurobiological critical period for the development of higher-order cognition. Neuroscience and Biobehavioral Reviews 94 (June), 179–195. https://doi.org/10.1016/j.neubiorev.2018.09.005.

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., Palminteri, S., 2017. Behavioural and neural characterization of optimistic reinforcement learning. Nature Human Behaviour 1 (4), 1–9. https://doi.org/10.1038/s41562-017-0067.

Luna, B., 2009. Developmental Changes in Cognitive Control through Adolescence. Adv Child Dev Behav 37, 233–278. https://doi.org/10.1146/annurev.biochem.78.082307.091526.Regulation.

McAuley, T., White, D.A., 2011. A latent variables examination of processing speed, response inhibition, and working memory during typical development. Journal of Experimental Child Psychology 108 (3), 453–468. https://doi.org/10.1016/j.jecp.2010.08.009.

Miller, E.K., Cohen, J.D., 2001. An Integrative Theory Of Prefrontal Cortex Function, pp. 167–202.

Montez, D.F., Calabro, F.J., Luna, B., 2017. The expression of established cognitive brain states stabilizes with working memory development. ELife 6, 1–26. https://doi.org/10.7554/eLife.25606.

Montez, D.F., Calabro, F.J., Luna, B., 2019. Working memory improves developmentally as neural processes stabilize. PLoS ONE 14 (3), 1–15. https://doi.org/10.1371/journal.pone.0213010.

Murty, V.P., Calabro, F., Luna, B., 2016. The role of experience in adolescent cognitive development: Integration of executive, memory, and mesolimbic systems. Neuroscience and Biobehavioral Reviews 70, 46–58. https://doi.org/10.1016/j.neubiorev.2016.07.034.

Myers, C., Shohamy, D., Clark, J., Gluck, M.A., Paré-Blagoev, E.J., Creso Moyano, J., Poldrack, R.A., 2002. Interactive memory systems in the human brain. Nature 414 (6863), 546–550. https://doi.org/10.1038/35107080.

Nagy, Z., Westerberg, H., Klingberg, T., 2004. Maturation of white matter is associated with the development of cognitive functions during childhood. Journal of Cognitive Neuroscience 16 (7), 1227–1233. https://doi.org/10.1162/0898929041920441.

Nassar, M.R., Frank, M.J., 2016. Taming the beast: Extracting generalizable knowledge from computational models of cognition. Current Opinion in Behavioral Sciences 11, 49–54. https://doi.org/10.1016/j.cobeha.2016.04.003.

Noble, K.G., McCandliss, B.D., Farah, M.J., 2007. Socioeconomic gradients predict individual differences in neurocognitive abilities. Developmental Science 10 (4), 464–480. https://doi.org/10.1111/j.1467-7687.2007.00600.x.

Noble, K.G., Norman, M.F., Farah, M.J., 2005. Neurocognitive correlates of socioeconomic status in kindergarten children. Developmental Science 8 (1), 74–87. https://doi.org/10.1111/j.1467-7687.2005.00394.x.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G.D.A., Conway, A., Cowan, N., Ward, G., 2018. Benchmarks of models of short-term and working memory. Psychological Bulletin. https://doi.org/10.5962/bhl.title.7369.

Oshri, A., Hallowell, E., Liu, S., MacKillop, J., Galvan, A., Kogan, S.M., Sweet, L.H., 2019. Socioeconomic hardship and delayed reward discounting: Associations with working memory and emotional reactivity. Developmental Cognitive Neuroscience 37, 100642. https://doi.org/10.1016/J.DCN.2019.100642.

Palminteri, S., Kilford, E.J., Coricelli, G., Blakemore, S.J., 2016. The Computational Development of Reinforcement Learning during Adolescence. PLoS Computational Biology 12 (6), 1–25. https://doi.org/10.1371/journal.pcbi.1004953.

Palminteri, S., Wyart, V., Koechlin, E., 2017. The Importance of Falsification in Computational Cognitive Modeling. Trends in Cognitive Sciences 21 (6), 425–433. https://doi.org/10.1016/j.tics.2017.03.011.

Peper, J.S., Hulshoff Pol, H.E., Crone, E.A., van Honk, J., 2011. Sex steroids and brain structure in pubertal boys and girls: A mini-review of neuroimaging studies. Neuroscience 191, 28–37. https://doi.org/10.1016/j.neuroscience.2011.02.014.

Petersen, C., Crockett, L., 1988. <Puberty Scale.Pdf> (2). https://doi.org/10.1103/PhysRevLett.78.1396.

Piekarski, D.J., Boivin, J.R., Wilbrecht, L., 2017. Ovarian Hormones Organize the Maturation of Inhibitory Neurotransmission in the Frontal Cortex at Puberty Onset in Female Mice. Current Biology 27 (12), 1735–1745. https://doi.org/10.1016/j.cub.2017.05.027 e3.

Piekarski, D.J., Johnson, C.M., Boivin, J.R., Thomas, A.W., Lin, W.C., Delevich, K., Wilbrecht, L., 2017. Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? Brain Research 1654, 123–144. https://doi.org/10.1016/j.brainres.2016.08.042.

Potter, T.C.S., Bryce, N.V., Hartley, C.A., 2017. Cognitive components underpinning the development of model-based learning. Developmental Cognitive Neuroscience 25, 272–280. https://doi.org/10.1016/j.dcn.2016.10.005.

Rosa, M.J., Bestmann, S., Harrison, L., Penny, W., 2010. Bayesian model selection maps for group studies. NeuroImage 49 (1), 217–224. https://doi.org/10.1016/j.neuroimage.2009.08.051.

Schulz, E., Wu, C., Ruggeri, A., & Meder, B. (n.d.). Searching for rewards like a child means less generalization and more directed exploration. BioRxiv.

Selmeczy, D., Fandakova, Y., Grimm, K.J., Bunge, S.A., Ghetti, S., 2018. Longitudinal trajectories of hippocampal and prefrontal contributions to episodic retrieval: Effects of age and puberty. Developmental Cognitive Neuroscience. https://doi.org/10.1016/j.dcn.2018.10.003 (June), 100599.

Sharpe, M.J., Chang, C.Y., Liu, M.A., Batchelor, H.M., Mueller, L.E., Jones, J.L., Schoenbaum, G., 2017. Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nature Neuroscience 20 (5), 735–742. https://doi.org/10.1038/nn.4538.

Shulman, E.P., Smith, A.R., Silva, K., Icenogle, G., Duell, N., Chein, J., Steinberg, L., 2016. The dual systems model: Review, reappraisal, and reaffirmation. Developmental Cognitive Neuroscience 17, 103–117. https://doi.org/10.1016/J.DCN.2015.12.010.

Somerville, L.H., Casey, B.J., 2010. Developmental neurobiology of cognitive control and motivational systems. Current Opinion in Neurobiology 20 (2), 236–241. https://doi.org/10.1016/j.conb.2010.01.006.

Somerville, L.H., Jones, R.M., Casey, B.J., 2010. A time of change: Behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. Brain and Cognition 72 (1), 124–133. https://doi.org/10.1016/j.bandc.2009.07.003.

Somerville, L.H., Sasse, S.F., Garrad, M.C., Drysdale, A.T., Akar, N.A., Insel, C., Wilson, R.C., 2017. Charting the expansion of strategic exploratory behavior during adolescence. Journal of Experimental Psychology: General 146 (2), 155–164. https://doi.org/10.1037/xge0000250.

Steinberg, L., 2005. Cognitive and affective development in adolescence. Trends in Cognitive Sciences 9 (2), 69–74. https://doi.org/10.1016/j.tics.2004.12.005.

Sutton, R.S., Barto, A.G., 2017. Reinforcement Learning: An introduction, 2nd ed. MIT Press.

Van Den Bos, W., Cohen, M.X., Kahnt, T., Crone, E.A., 2012. Striatum-medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. Cerebral Cortex 22 (6), 1247–1255. https://doi.org/10.1093/cercor/bhr198.

van den Bos, W., Güroğlu, B., van den Bulk, B.G., Rombouts, S.A.R.B., Crone, E.A., 2009. Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. Frontiers in Human Neuroscience 3 (December), 1–11. https://doi.org/10.3389/neuro.09.052.2009.

Wilson, R., Collins, A.G.E., 2019. Ten simple rules for the computational modeling of behavioral data. PsyArXiv. In press.

Wimmer, G.E., Braun, E.K., Daw, N.D., Shohamy, D., 2014. Episodic Memory Encoding Interferes with Reward Learning and Decreases Striatal Prediction Errors. Journal of Neuroscience 34 (45), 14901–14912. https://doi.org/10.1523/jneurosci.0204-14.2014.